KEY PROGRAM MEASUREMENT CONCEPTS AND DEFINITIONS

The "family of measures" is a technique for providing a comprehensive overview of the performance of a program or program element from multiple perspectives – ranging from what it has cost (inputs) to what it has achieved (outcomes) and how efficient it has been in producing those results – while serving the needs of managers, staff, senior County management, elected officials, and the public. In this appendix, we introduce the key concepts and definitions needed to understand the family of measures.

The definitions presented here are consistent with those proposed by the Montgomery County Council in Resolution 13-1488.[1] A few have been adapted or elaborated to make them more compatible with the overall *Montgomery Measures Up!* framework and the strategies and distinctions emphasized there. In some cases, the definitions provided here are somewhat narrower than those used by other jurisdictions. This has been done to simplify the synthesis of the measures and to ensure consistency with the definitions given in Council Resolution 13-1488.

The following section provides a brief conceptual introduction to the notion of performance measurement, the family of measures, and the definitions used in *Montgomery Measures Up!* It is followed by a review of important relationships between the various types of program and community outcome measures – and the ambiguities that can arise concerning how to classify a particular measure. The appendix closes with a discussion of the benefits of using the *family* of measures approach.

Program Measures and the Causal Chain

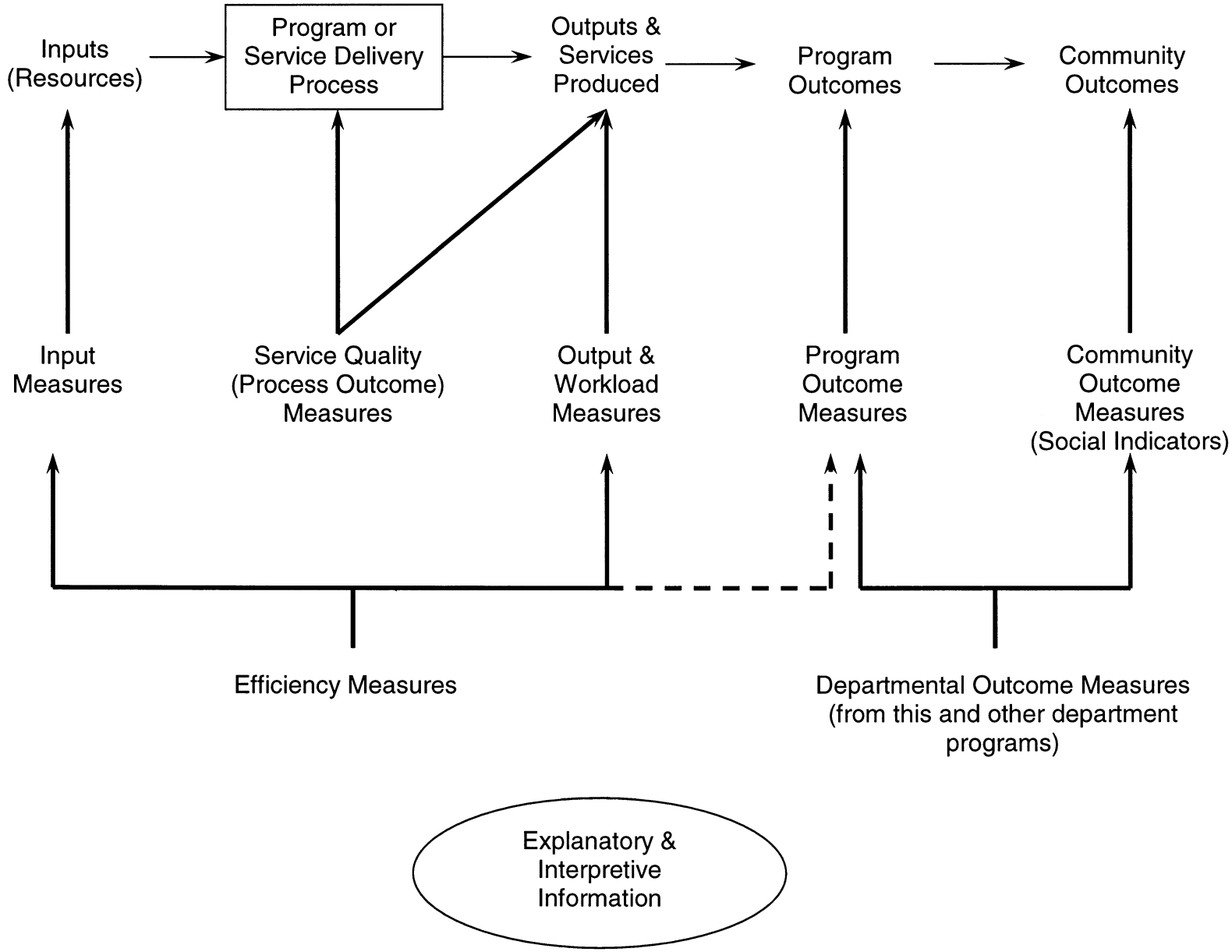
It is perhaps easiest to understand the many types of program measures, their distinctions, and their interrelations by focusing on the concept of a "causal chain."[2] The various types of performance measures can be viewed as a *spectrum* of information. This can be seen most easily by considering how, in general terms, a program or service operates and the information it produces in affecting clients, users, and other citizens or target populations (see Exhibit 1).

---

[1] These were, in turn, adapted from the definitions adopted by the Governmental Accounting Standards Board, the State of Maryland, and Fairfax County, Virginia.

[2] See, in particular, James E. Swiss, *Public Management Systems: Monitoring and Managing Government Performance*, Prentice-Hall (Englewood Cliffs, New Jersey, 1991), chapter 5; *Measuring Programs – A Practical Approach*, United Way of America (Alexandria, Virginia, 1996); and John M. Greiner, "Positioning Performance Measurement for the Twenty-First Century," in *Organizational Performance and Measurement in the Public Sector*, ed. Arie Halachmi and Geert Bouckaert, Quorum Books (Westport, Connecticut, 1996), chapter 1.

Exhibit 1



RELATIONSHIPS BETWEEN THE VARIOUS PROGRAM MEASURES

A government program or service can be viewed as a process (some would call it an "intervention") whose effects ripple through the county and its citizens in ever widening circles. This service delivery process combines various resources – employees, materials, equipment, funding, management skills, etc. – and converts them into products and services that are designed to address a particular need: provide recreation, assist the homeless, provide computer support for County employees, etc. All programs and services produce a chain of effects, beginning with direct effects on the citizens or organizations seeking the service and proceeding to ever-more indirect effects, some of them emerging only after numerous intervening processes have come into play. A causal chain is a rough model (in a sense, a *theory*) of how a program converts inputs into immediate outputs or events whose widening impacts are expected, ultimately, to address the reason for providing the service or product in the first place. Those impacts, of course, do not usually stop after addressing the immediate reason for producing the service. Usually they continue to ripple outward, ultimately affecting, if ever so slightly, the fundamental goals and concerns of County citizens – quality educational opportunities, safety and public order, economic security and opportunity, quality of life, etc.

A causal chain is simply a diagram of the expected outputs and their successive impacts for a single program or program element. One way to generate such a chain is to start with the immediate result (output) of the program and ask, "Why do we want this product or service?" Take the answer and ask the same question again – "Why do we want this result?" By repeating the process, one will generate a lengthening chain of successive causes and effects that should (if the program does in fact affect the problem it is designed to address!) encompass the reason for providing the program in the first place. And by continuing to ask such questions, one will ultimately be led to fundamental philosophical/community values such as those mentioned above.

For example, suppose the program of interest consists of classes offered through the Department of Recreation. The program (service delivery process) involves activities such as deciding on the course offerings, securing instructors and classroom space, scheduling the classes, publicizing/promoting the offerings, registering the participants, and providing instruction to those who attend. The major products of this program consist of classes offered and participants who sign up for the classes (partly or wholly because of the promotional activities of the department) and who attend some or all of the sessions. Suppose that one starts the "why" questions with the program activities – why offer and promote the classes? Answer: to get people to sign up for and participate in the courses. Why do we want them to take the courses? So that they can learn new skills, meet people, enjoy their leisure time, etc. And why do we want these things? To give citizens a chance to improve their education, boost their economic opportunities, increase their quality of life, possibly even enhance their safety and security (for instance, if they are taking self-defense courses, or if the courses help keep children in a secure environment after school).

Different people may arrive at somewhat different sequences – some longer, some shorter, some with more branches, some with fewer – but the pattern will generally be the same. For purposes of defining Montgomery County's family of measures, the causal

chain can be split into five key regions.[1] The top row of Exhibit 1 shows a general causal chain for a government program (or program element) and the five regions of interest. The chain consists of a sequence of causes and effects that starts with the inputs to the program and ends with very general "community" outcomes.

- *Inputs* are the resources needed for the given program to function: staff, office supplies, equipment, vehicles, office space, energy, managerial skills and time, funds for contractors, etc.

- The *program or service delivery process* corresponds to the activities that together allow the various inputs to be converted into the intended products or services. This is what government *does* – activities such as responding to customer requests, reviewing permits, patrolling streets, or inspecting food establishments – all carried out in pursuit of a program's mission. Note that the program or service delivery process does not have to focus on customers or organizations *external* to the government. Internal programs or services such as procurement, accounting, and computer services can also be considered using this model.

- *Outputs* are the products and services produced by the program or service delivery process. These are the most immediate results of the program – the tangible effects of delivering the service. For example, clients advised, fires extinguished, buildings inspected, complaints addressed, or meals delivered and consumed could all be the outputs of government programs.

- *Program outcomes* are the immediate changes or benefits experienced by the clients or other groups targeted by the program. (Again, these clients may be external to the government or internal government users.) Here, the focus is on the person or group directly affected by the program or service, rather than the impact on the community as a whole. Program outcomes are the effects most closely related to and influenced by a program's outputs (*and* by day-to-day management of the program). Often such outcomes are not ends in themselves but are expected to lead to – or facilitate – achievement of the desired end result (the program's mission). When the desired result (such as *stable* employment for a homeless person) cannot be assured until considerable time has elapsed, an intermediate program outcome (in this case, getting a job) is especially important from the standpoint of managing and assessing the program.

- *Community outcomes* involve broadly stated results from the perspective of the whole community. These represent desired outcomes for the community *as a whole* (rather than for individual clients) and are located toward the end of the causal chain. As such, they may be influenced by other programs housed in a

[1] While this model has been simplified and compressed in some respects (for instance, some advocate the use of additional "types" of outcomes), it serves to illustrate the origin – and importance – of the various items in the family of measures and the relationships between the many indicators that are the focus of *Montgomery Measures Up!*

given department, or by programs from several different departments, as well as private sector activities. They are also highly subject to factors outside the control of these groups – demographic changes, climate and geography, economic conditions, etc. Thus the ability of any one program or program element to influence community outcomes is usually limited. Nevertheless, such outcomes are often of great interest to the public and need to be included in any comprehensive picture of County performance.

The various processes, events, and effects along the causal chain produce (or *could* produce, if measured) various types of information, what some have called a "chain of evidence."[1] Thus, at each stage of the causal chain, one can define a group of measures describing that stage. This is illustrated by the second row in Exhibit 1.

- *Input measures* describe the resources such as dollars or workyears used to provide the product or service associated with the program.

- *Output measures* quantify the direct products of the program or service delivery process – volume of work accomplished, units of service provided, number of items produced, etc. Examples include clients counseled, applications received, lane-miles of road resurfaced, telephone "troubles" repaired, and GIS maps produced.

- *Service quality measures* are a special type of program outcome that address the merit of the *program or service delivery process **itself*** – for instance, its accuracy, its timeliness, and customer satisfaction with the way services were provided. Other important aspects of service quality include the accessibility of the service, the degree of crowding at the service facility, and the courtesy and helpfulness of service staff. Thus, service quality reflects the customer's expectations and the degree to which they are met. As Exhibit 1 suggests, these measures usually involve information about the service delivery process itself as well at the outputs produced (e.g. their accuracy). Service quality is a good example of an intermediate outcome in that while it is clearly important (especially since customer satisfaction and responsive government are explicit elements of the County's Vision Statement and Guiding Principles), it is unlikely to represent the ultimate mission of a given program or service.

- *Program outcome measures* characterize the immediate *results* of the program's outputs on those receiving the services – the impact on the program's clients, users, or other target group. In contrast to service quality measures, program outcome measures refer to benefits accruing to participants after service delivery. Examples include lives saved, skills learned, fires extinguished, diseases prevented, (reduced) child abuse cases reopened within six months, (increased) recycling behavior, etc. Program outcomes also include negative or undesired

[1] U.S. General Accounting Office, *Measuring Performance and Demonstrating Results of Information Technology Investments*, Accounting and Information Management Division, Report GAO/AIMD-98-89 (U.S. General Accounting Office, Washington D.C., March 1998), p. 26.

results that may emerge in connection with a program, such as recidivism by persons completing drug treatment or the rate of breakdowns for County-repaired vehicles.[1]

- *Community outcome measures* characterize general conditions of well-being for an entire population or community, whereas program outcome measures focus on the individual clients and target groups directly affected by a given program. The focus of community outcome measures is on broad social indicators related to the fundamental goals and values of the County and its citizens – public safety, quality of life, excellence in education, healthy children and adults, young people making smart choices, etc. Community outcome measures often involve statistics from special censuses and similar sources that are able to provide information about the entire community. Because such indicators usually reflect the results of numerous programs – both public and private – as well as powerful economic, social, and other external forces, it is inappropriate to imply that they are the direct or primary responsibility of a single program or program element. Indeed, such outcomes may go beyond the original mission and scope of the program. To be accurate, the most that one can often say is that a given program or program element *supports* certain community outcomes.

Note that in some cases, it is not practical to measure certain key program (or community) outcomes of interest. For instance, once the customer leaves the service area, it may be impractical to follow up on his or her behavior in response to receiving the service. Sometimes the program results require a long time for the effects to become apparent. (For instance, the achievement of *stable* employment for disadvantaged youth can, by definition, only be assessed after several months or years of job experience have been accumulated.) In other cases, measurement of the program outcome of interest may be too costly, technically difficult, or intrusive to be practical. For example, it is impractical to assess the impact of home-delivered or congregate meals on the health and nutrition of the senior citizens receiving those meals.

In such cases, it is usually necessary to use a more practical "proxy" measure to assess the outcomes. Attendance, customer satisfaction, hours of "positive" recreation activities provided, number of children vaccinated for measles, and the number of in-home and congregate meals provided to seniors can all be considered proxy measures of the outcomes of County programs. "Proxy measures" are surrogates or stand-ins for the actual outcome one wants to measure. Proxies are *assumed* to be highly correlated with the outcome of interest, so that a trend in a proxy is likely to correspond to the trend that would be observed if one could actually measure the program outcome of interest.

---

[1] Many outcome measures can be written to focus on either the positive or the negative aspects of a program's results. For instance, one can define an outcome measure for a drug treatment program as "the percentage of addiction treatment clients remaining drug-free for one year" or, equivalently, "the percentage of such clients whose cases were reopened within a year." While many managers may prefer the former statement of the measure for reporting results to the public, from a *management* perspective the negative form of the measure is probably much more significant and useful for monitoring (and maintaining or improving) program effectiveness.

When there is a strong correlation between a given proxy measure and the outcome of interest, such a measure can be an excellent substitute. In such instances, the links in the causal chain are so clear and reliable that the occurrence of the output almost certainly leads to the desired program outcome. For instance, the number of children vaccinated for measles and the number of seniors served home-delivered or congregate meals are usually excellent proxies for the effectiveness of those respective programs.[1]

However, to the extent that the correlation between a proxy measure and the program outcome of interest is weaker, the proxy is less useful. The value of a given proxy measure depends critically upon the assumptions that must be made in linking it to the outcome of interest and the likelihood that they will be violated. Of course, sometimes a proxy measure is the best that one can do, especially until better data collection procedures can be developed. And even poor proxy measures can be useful in alerting managers and the public to *potential* problems with the effectiveness of a program. Nevertheless, one must be cautious when utilizing proxy measures of program and community outcomes.

Two other types of information that can be related to the causal chain – efficiency measures and departmental outcome measures – are also of interest in connection with the family of measures approach. As shown in Exhibit 1, these two types of measures are largely derived from other measures on the causal chain.

- *Efficiency measures* depict the relationship between the amount of resources used and the amount of work performed or services produced. They indicate how well the program is using resources to produce goods and services. Usually this is shown as the ratio of two types of measures, inputs and outputs – the cost per unit output, the number of units of output produced per dollar or per workyear, etc. Examples include the average cost of a child abuse and neglect investigation, the average cost per GIS map produced, the number of library books circulated per workyear, the number of plans reviewed per reviewer, and the cost per criminal investigation. Another (less common) type of efficiency measure combines inputs and *outcomes* to demonstrate how much it costs to produce a given program outcome (e.g. the cost per criminal investigation resulting in an arrest). Staff utilization percentages are also sometimes employed as efficiency indicators. Efficiency measures have the advantage of being more controllable by management than outcome measures and being readily understood by non-program personnel.

- *Departmental outcome measures* reflect the degree to which the overall mission of the department is being achieved. Such measures are usually drawn from the

[1] Children vaccinated against measles can be reliably expected *not* to contract measles, the mission of the vaccination program. In the case of meals served to senior citizens, the presumption is that eating those meals will maintain or improve the health and nutrition of the seniors attending – the mission of the program. However, if "meals served to seniors" is to be used as a proxy program outcome, one must be confident that those meals are nutritious, sufficiently appetizing to actually be consumed, and appropriate to the dietary needs of the seniors being fed. While probably a good assumption in most cases, this should be verified from time to time if "meals served to seniors" continues to be used as a proxy outcome measure.

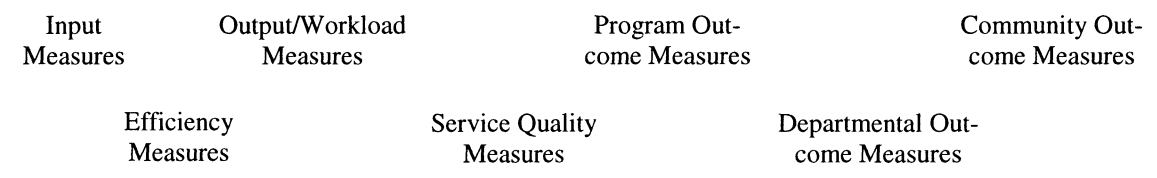
program outcome measures for the organization's major programs, as well as certain community outcomes that are especially salient for the given department (see Exhibit 1). Departmental outcomes often tend to be closer to community outcomes than to program outcomes on the causal chain, although sometimes a program so dominates a department's efforts that the corresponding program outcomes are, to all intents and purposes, the departmental outcomes. Several departments noted that the family of measures' focus on individual program outcomes is often too fragmented or incomplete to capture the breadth and scope of the department's mission and how well it is being achieved.[1] Indeed, sometimes a key *departmental* outcome does not appear as a specific outcome for any of its individual programs but instead emerges from the combined effects of several programs. Examples include the crime rate (for the Police Department) or ambient air and water pollution levels (for the Department of Environmental Protection). If one only shows individual program outcomes as part of the family of measures, the broader departmental outcomes may slip through the cracks, leading to an incomplete (and possibly distorted) picture of departmental performance. This year departmental outcomes have been reported by eight departments, including Environmental Protection, Health and Human Services, and the Fire and Rescue Service.

Finally, a depiction of program performance is often not complete without including some explanatory and interpretive information. Such information can provide the context for the results shown in the family of measures and give the reader an understanding of what the figures mean. A description of special initiatives being undertaken by the department (either to address current deficiencies or to enhance future results) and information on external factors that have affected (or are expected to affect) the program are also relevant here. While not linked to a specific region of the causal chain, such information can be very important to understanding the results for a given program.

Important Relationships Between the Measures

It is useful to consider the various types of performance measures according to where they are located on the causal chain. Exhibit 2 depicts such an arrangement. (Since service quality measures are, in effect, a type of outcome, it is more appropriate to group them closer to program outcomes than is shown in Exhibit 1.)

Exhibit 2



[1] This is especially true early in the process of implementing the family of measures, when most departments are only reporting on a few of their major programs.

As one moves from left to right in Exhibit 2 (e.g. along the causal chain), a number of (sometimes contradictory) trends are exhibited by the various types of measures:

MOVING TO THE LEFT IN EXHIBIT 2	MOVING TO THE RIGHT IN EXHIBIT 2
Results are more concrete: tangible, easier to measure	Results are less concrete and harder to measure
Results/information are more immediate and more amenable to daily or monthly reporting	Results take longer to materialize, and frequent monitoring and reporting are more difficult
More directly affected by the service or program itself than by external factors	Affected to a greater extent by other programs, organizations, and external factors
More amenable to being used for day-to-day management	More amenable for use in periodic planning, budgeting, and policy decisions
Measures are more reliable (accurate and repeatable)	Measures tend to be less reliable and more subject to measurement error
Measures are less valid as indicators of the program's (or department's) overall effectiveness (the degree to which it is achieving its mission)	Measures are more valid indicators of the extent to which a program or department is effective in achieving its mission
Results are more controllable by management	Results are less controllable by management

The regions between different types of measures (e.g. between input and workload measures, output and program outcome measures, service quality and program outcome measures, efficiency and service quality measures) often give rise to confusion in terms of how to classify a measure. Is the number of clients arriving at a clinic an input or a workload/output measure? Is attendance at recreation classes an output or a program outcome? What about children vaccinated at a County clinic – is this an output or a program outcome? Is police response time an outcome or a service quality measure, or is it an indication of efficiency? We have attempted to provide some guidance (e.g., conventions) for such cases,[1] but as explained below, the family of measures format reduces the importance (and frustration!) associated with how one classifies any given measure.

[1] "A Primer on Program Measures for Montgomery County," Office of Management and Budget, Montgomery County Government (Rockville, Maryland, December 15, 1999).

Why Use a *Family* of Measures?

The need to report on an *entire* family of measures to characterize the performance of a given program or program element may seem excessively burdensome when it would appear that merely providing a few quality, efficiency, or outcome measures would suffice. But as shown above, each type of performance measure has its own advantages and disadvantages. This arises in part because the links in the causal chain for a government program are not cleanly delineated with well-defined causes and results. Running a government program is not like turning on a light or building a car: it is not just a matter of adjusting the settings on a machine or assembly line, following a prescribed procedure, and obtaining the desired result. The links in the causal chain for a government program are often fuzzy, uncertain, and affected by external factors, some of which are only dimly perceived at the time. A program's outputs cannot be accurately predicted just by knowing its inputs, nor can one infer the outcomes or service quality associated with a government program just by reporting its outputs. **To understand and effectively manage government programs, one must monitor *each* type of measure along the causal chain**. Indeed, given the diverse audience for information on the performance of such programs, each type of measure has its advocates and advantages. (Management tends to value the more controllable measures to the left of Exhibit 2, while elected officials and the general public tend to be more interested in the measures to the right of Exhibit 2, plus any efficiency measures.)

The family of measures initiative has been designed as the first step in a long-term effort to provide a comprehensive picture of County performance. A "family" of measures involves **the measurement of organizational or program performance with a set of measures consisting of inputs, outputs, service quality indicators, outcomes, and efficiency measures *at the program level.*** Thus, the family of measures describes and quantifies the first part of the causal chain for a program or program element – all but the community outcomes. For the time being, community outcome measures – an especially difficult and sometimes costly aspect of performance measurement – will be addressed only indirectly by noting the community outcomes *supported* by a given program or program element.[1]

A number of benefits accrue from using families of measures as defined above:

- The family of measures presents a balanced picture of program performance. Prior to its introduction, the emphasis was largely on inputs (dollars and positions) and, to a lesser extent, outputs.
- The family as defined above emphasizes the measures that are of greatest value to managers (e.g. the left portion of the causal chain).

[1] However, note that the Department of Health and Human Services has made considerable progress in assessing community outcomes as part of its "Measuring Progress" reports and the Children's Agenda. (See *Measuring Progress Report 1998*, Department of Health and Human Services; *Measuring Progress: A Strategy to Get Results*, Department of Health and Human Services, June 1999, June 2000, and July 2001; and *The Children's Agenda*, Montgomery County Collaboration Council for Children, Youth & Families.)

- At the same time, because the causal links in government programs are often fuzzy, uncertain, and influenced by external factors and events, the family of measures approach recognizes the importance of including information from *each* region of the causal chain. Managers need many types of information to effectively oversee public sector programs.
- The presence of multiple measures and multiple types of measures for each program or program element tends to balance and in some cases compensate for the biases and idiosyncrasies of particular measures. Thus, there is a place in the family of measures for both the easy and the more difficult measures, for those measures that are very reliable and those that may be more prone to random error, for those that are highly controllable by managers and those that are less controllable, for those that are of primary interest to management and those on which the public tends to focus. Since *all* departments reporting program measures are supposed to provide each type of measure, organizations are less concerned about displaying a few measures over which they have limited control or which are prone to high levels of uncertainty or large fluctuations.
- Using a *set* of measures can guard against the tendency to focus on one measure or type of measure to the exclusion of other important aspects of performance – e.g. emphasizing outputs while downplaying outcomes, or sacrificing service quality in pursuit of efficiency.
- And finally, using the family of measures reduces the tendency to quibble over how to classify a given program measure. People (and governments) will sometimes differ on how to label a given measure. However, an important advantage to using a *family* of measures is that regardless of where a particular measure is finally entered on the program measures display, the family of measures format ensures a complete picture of the performance of the given program or program element. Whether one happens to classify an ambiguous item as an output, outcome, efficiency, or service quality measure is less important to most users than the fact that the information is relevant, important, accurate, and included *somewhere* on the program measures display.

In a sense, using a family of measures allows one to "triangulate" the performance of the given program or program element. Program performance is presented using a variety of complementary measures and perspectives: what it has cost (inputs), what it has achieved (outcomes), and how efficient it has been in producing those results. Taken together (possibly with the help of explanatory information), the family of measures allows one to formulate a reasonably complete picture of a program and its performance.